

CLAIMS

What is claimed is:

1. A machine implemented method for deriving a sequence of at least a portion of an oligomer from a mass spectrum data of fragments of said oligomer, said method comprising:
 - providing a predetermined set of mass/charge (m/z) values for monomer sequences;
 - determining an abundance value from said mass spectrum data for each m/z value in said predetermined set, thereby producing a plurality of abundance values;
 - calculating a first ranking, based on said plurality of abundance values, for each sequence of a set of fragment sequences having a first number of monomers;
 - calculating a second ranking, based on said plurality of abundance values, for each sequence of a set of fragment sequences having a second number of monomers;
 - calculating a cumulative ranking, based on said first ranking and said second ranking, for each sequence of a set of fragment sequences having at least said second number of monomers.
2. A method as in claim 1 wherein said oligomer is a protein.
3. A method as in claim 2 wherein said portion of said protein is a terminal portion of said protein.
4. A method as in claim 3 wherein said terminal portion is one of an N-terminus or a C-terminus.
5. A method as in claim 3 wherein a label is attached to said portion.

6. A method as in claim 5 wherein said label is covalently bonded to said protein prior to generating said mass spectrum data and wherein said mass spectrum data is transformed from an output of a detector plate.
7. A method as in claim 6 wherein said protein is fragmented by collision-induced dissociation to generate fragments, comprising said portion, which are then accelerated toward a detector plate to generate said mass spectrum data.
8. A method as in claim 3 wherein said protein is fragmented to generate fragments, comprising said portion, where are then accelerated toward a detector plate to generate said mass spectrum data.
9. A method as in claim 2 wherein said protein is isolated from other proteins extracted from a sample and wherein said machine which implements said method comprises a digital processing system which executes computer programming instructions.
10. A method as in claim 3 wherein said predetermined set comprises all possible m/z values empirically found in mass spectra for all possible amino acid sequences having a number of amino acids from one amino acid to a selected number of amino acids, said selected number being in a range from 4 to 8 amino acids.
11. A method as in claim 2 wherein said predetermined set comprises, for a given sequence of a given number of amino acids, a set of fragment types and a set of ionic charge states.
12. A method as in claim 2 wherein said set of amino acid sequences having a first number of amino acids and said set of amino acid sequences having a second number of

amino acids comprise all possible amino acid sequences for both said first number of amino acids and said second number of amino acids.

13. A machine implemented method for deriving a sequence of at least a portion of an oligomer from a mass spectrum data, said method comprising:

providing a predetermined set of mass/charge (m/z) values for monomer sequences each of which comprises a mass label;

determining an abundance value from said mass spectrum data for each m/z value in said predetermined set, thereby producing a plurality of abundance values;

calculating a first ranking, based on said plurality of abundance values, for each sequence of a set of monomer sequences having a first number of monomers.

14. A method as in claim 13 wherein said oligomer is a protein.

15. A method as in claim 13 wherein said mass label has a mass which is different than a mass of each possible amino acid in said set of amino acid sequences.

16. A method as in claim 13 wherein said mass label imparts a unique mass signature to each sequence of said set of amino acid sequences.

17. A method as in claim 13 wherein said portion is a terminal portion of said protein.

18. A method as in claim 17 wherein said terminal portion is one of an N-terminus or a C-terminus.

19. A method as in claim 18 wherein said mass label is covalently bonded to said terminal portion prior to generating said mass spectrum data.

20. A method as in claim 19 wherein said protein is fragmented in a mass spectrometer to generate fragments, comprising said portion, which are then accelerated toward a detector plate to generate said mass spectrum data.

21. A method as in claim 20 wherein said protein is isolated from other proteins extracted from a sample and wherein said machine which implements said method comprises a digital processing system which executes computer programming instructions.

22. A method as in claim 14 wherein said predetermined set comprises all possible m/z values empirically found in mass spectra for all possible amino acid sequences, each of which comprises a mass label, having a number of amino acids from one amino acid to a selected number of amino acids, said selected number being in a range from 4 to 8 amino acids.

23. A method as in claim 14 wherein said predetermined set comprises, for a given sequence of a given number of amino acids, a set of fragment types and a set of ionic charge states.

24. A method as in claim 14 wherein said set of amino acid sequences having a first number of amino acids comprises all possible amino acid sequences for said first number of amino acids.

25. A method as in claim 2 wherein said method is performed for each protein in a set of proteins extracted from a biological material and wherein said set of proteins is more than 100 different proteins.

26. A method as in claim 14 wherein said method is performed for each protein in a set of proteins extracted from a biological material and wherein said set of proteins is more than 100 different proteins.

27. A machine readable medium containing executable computer program instructions, which when executed by a processing system cause said processing system to perform a method for deriving a sequence of at least a portion of an oligomer from a mass spectrum data, said method comprising:

providing a predetermined set of mass/charge (m/z) values for monomer sequences;
determining an abundance value from said mass spectrum data for each m/z value in said predetermined set, thereby producing a plurality of abundance values;
calculating a first ranking, based on said plurality of abundance values, for each sequence of a set of fragment sequences having a first number of monomers;
calculating a second ranking, based on said plurality of abundance values, for each sequence of a set of fragment sequences having a second number of monomers;
calculating a cumulative ranking, based on said first ranking and said second ranking, for each sequence of a set of fragment sequences having at least said second number of monomers.

28. A machine readable medium as in claim 27 wherein said oligomer is a protein.

29. A machine readable medium as in claim 28 wherein said portion of said protein is a terminal portion of said protein.

30. A machine readable medium as in claim 29 wherein said terminal portion is one of an N-terminus or a C-terminus.

31. A machine readable medium as in claim 29 wherein a label is attached to said portion.
32. A machine readable medium as in claim 31 wherein said label is covalently bonded to said portion prior to generating said mass spectrum data.
33. A machine readable medium as in claim 32 wherein said protein is fragmented to generate fragments, comprising said portion, which are then accelerated toward a detector plate to generate said mass spectrum data.
34. A machine readable medium as in claim 29 wherein said protein is fragmented to generate fragments, comprising said portion, where are then accelerated toward a detector plate to generate said mass spectrum data.
35. A machine readable medium as in claim 28 wherein said protein is isolated from other proteins extracted from a sample and wherein said machine which implements said method comprises a digital processing system which executes computer programming instructions.
36. A machine readable medium as in claim 29 wherein said predetermined set comprises all possible m/z values empirically found in mass spectra for all possible amino acid sequences having a number of amino acids from one amino acid to a selected number of amino acids, said selected number being in a range from 4 to 8 amino acids.
37. A machine readable medium as in claim 28 wherein said predetermined set comprises, for a given sequence of a given number of amino acids, a set of fragment types and a set of ionic charge states.

38. A machine readable medium as in claim 28 wherein said set of amino acid sequences having a first number of amino acids and said set of amino acid sequences having a second number of amino acids comprise all possible amino acid sequences for both said first number of amino acids and said second number of amino acids.

39. A machine readable medium containing executable computer program instructions, which when executed on a processing system cause said processing system to perform a method for deriving a sequence of at least a portion of an oligomer from a mass spectrum data, said method comprising:

providing a predetermined set of mass/charge (m/z) values for monomer sequences each of which comprises a mass label;

determining an abundance value from said mass spectrum data for each m/z value in said predetermined set, thereby producing a plurality of abundance values;

calculating a first ranking, based on said plurality of abundance values, for each sequence of a set of monomer sequences having a first number of monomers.

40. A machine readable medium as in claim 39 wherein said oligomer is a protein.

41. A machine readable medium as in claim 40 wherein said mass label has a mass which is different than a mass of each possible amino acid in said set of amino acid sequences.

42. A machine readable medium as in claim 40 wherein said mass label imparts a unique mass signature to each sequence of said set of amino acid sequences.

43. A machine readable medium as in claim 40 wherein said portion is a terminal portion of said protein.

44. A machine readable medium as in claim 43 wherein said terminal portion is one of an N-terminus or a C-terminus.

45. A machine readable medium as in claim 44 wherein said mass label is covalently bonded to said terminal portion prior to generating said mass spectrum data.

46. A machine readable medium as in claim 45 wherein said protein is fragmented in a mass spectrometer to generate fragments, comprising said portion, which are then accelerated toward a detector plate to generate said mass spectrum data.

47. A machine readable medium as in claim 46 wherein said protein is isolated from other proteins extracted from a sample and wherein said machine which implements said method comprises a digital processing system which executes computer programming instructions.

48. A machine readable medium as in claim 40 wherein said predetermined set comprises all possible m/z values empirically found in mass spectra for all possible amino acid sequences, each of which comprises a mass label, having a number of amino acids from one amino acid to a selected number of amino acids, said selected number being in a range from 4 to 8 amino acids.

49. A machine readable medium as in claim 40 wherein said predetermined set comprises, for a given sequence of a given number of amino acids, a set of fragment types and a set of ionic charge states.

50. A machine readable medium as in claim 40 wherein said set of amino acid sequences having a first number of amino acids comprises all possible amino acid sequences for said first number of amino acids.

51. A machine readable medium as in claim 28 wherein said method is performed for each protein in a set of proteins extracted from a biological material and wherein said set of proteins is more than 100 different proteins.

52. A machine readable medium as in claim 40 wherein said method is performed for each protein in a set of proteins extracted from a biological material and wherein said set of proteins is more than 100 different proteins.

53. A method for processing noise in a mass spectrum data of a fragmented oligomer, said method comprising:
determining a substantially periodic block of noise in a mass spectrum data generated from accelerating fragments of an oligomer to a detector;
filtering said substantially periodic block of noise from said mass spectrum data.

54. A method as in claim 53 wherein said oligomer is a protein.

55. A method as in claim 54 wherein said protein is randomly fragmented with collision induced dissociation.

56. A method as in claim 54 wherein a mass label is attached to said protein on a terminal portion of said protein.

57. A method as in claim 56 wherein said protein is fragmented after said mass label is attached to said terminal portion.

58. A method as in claim 57 wherein said mass spectrum data is obtained from an in source mass spectrometer and wherein said protein is randomly fragmented in said in source mass spectrometer.

59. A method as in claim 57 wherein said substantially periodic block of noise is substantially independent of time.

60. A method as in claim 57 wherein said mass spectrum data, after said filtering, is used to identify an amino acid sequence of said protein.

61. A method as in claim 60 wherein said amino acid sequence is identified by determining an amino acid sequence of said terminal portion.

62. A method as in claim 57 wherein said mass label is covalently bonded to said terminal portion.

63. A method as in claim 57 wherein said protein is isolated from other proteins extracted from a biological sample and wherein a machine, which implements said method, comprises a digital processing system which executes computer programming instructions.

64. A method as in claim 63 wherein over 100 proteins are extracted from said biological sample and said method is performed for each of said 100 proteins.

65. A method as in claim 57 further comprising:

providing a predetermined set of mass/charge (m/z) values for amino acid sequences;

determining an abundance value from said mass spectrum data for each m/z value in said predetermined set, thereby producing a plurality of abundance values;

calculating a first ranking, based on said plurality of abundance values, for each sequence of a set of amino acid sequences having a first number of amino acids;

calculating a second ranking, based on said plurality of abundance values, for each sequence of a set of amino acid sequences having a second number of amino acids;

calculating a cumulative ranking, based on said first ranking and said second ranking, for each sequence of a set of amino acid sequences having at least said second number of amino acids.

66. A method as in claim 57 further comprising:

providing a predetermined set of mass/charge (m/z) values for amino acid sequences each of which comprises a mass label;

determining an abundance value from said mass spectrum data for each m/z value in said predetermined set, thereby producing a plurality of abundance values;

calculating a first ranking, based on said plurality of abundance values, for each sequence of a set of amino acid sequences having a first number of amino acids.

67. A method for determining a sequence of at least a portion of an oligomer from mass spectrum data, said method comprising:

reading mass spectrum data in a first reading operation from a non-volatile storage device to a temporary volatile cache memory to obtain abundance values at a set of possible mass/charge (m/z) values from said temporary volatile cache memory and calculating first abundance parameters from said abundance values;

reading said mass spectrum data in a second reading operation, following said first reading operation, from said temporary volatile cache memory to obtain said abundance values at said set of possible m/z values, and determining a ranking, based on said abundance values, for each sequence of a set of monomer sequences having a first number of monomers.

68. A method as in claim 67 wherein said oligomer is a protein.

69. A method as in claim 68 wherein said first abundance parameters and said ranking for said each sequence are stored in said temporary volatile cache memory.

70. A method as in claim 68 wherein said set of possible m/z values is calculated as needed rather than stored on said non-volatile storage device.

71. A method as in claim 69 wherein said ranking for said each sequence is determined from said first abundance parameters and said abundance values obtained in said second reading operation, and wherein said temporary volatile cache memory comprises at least one of an L1 and an L2 cache of a microprocessor.

72. A machine implemented method for determining a sequence of at least a portion of an oligomer from mass spectrum data, said method comprising:
determining a first molecular weight for a first monomer sequence;
determining a set of weight adjustments for possible ion types of said first monomer sequence;

determining a set of charge state adjustments for possible charge states of said possible ion types;

calculating a set of m/z values for said first monomer sequence from said first molecular weight, said set of weight adjustments and said set of charge state adjustments.

73. A method as in claim 72 wherein said oligomer is a protein.

74. A method as in claim 73 wherein said set of m/z values are used to perform lookup operations into a mass spectrum data to obtain abundance values and wherein said set of m/z values are not retained in a non-volatile storage device for access in an abundance value lookup operation.

75. A method as in claim 74 wherein said set of m/z values is stored in a temporary volatile cache memory when needed and is erased for subsequent lookup operations in said temporary volatile cache memory.

76. A method as in claim 75 wherein said mass spectrum data is stored in said temporary volatile cache.

77. A method as in claim 76 wherein said temporary volatile cache comprises at least one of an L1 or L2 cache of a microprocessor.

78. A method as in claim 1 wherein said mass spectrum is digitally filtered to minimize spectral noise prior to said determining said abundance value.

79. A method as in claim 1 wherein said providing of said predetermined set is one of (a) storing said predetermined set or (b) calculating needed portions of said predetermined set on an as-needed basis.

80. A method as in claim 1 wherein said protein is cleaved by collision induced dissociation, either in-source or in a collision cell to generate fragments which are then accelerated toward a detector plate.

81. A machine implemented method for deriving a sequence of at least the labeled terminal portion of a protein from a mass spectrum data, said method comprising:
labeling the protein with at least two labels that differ in mass from each other by at least 1 amu;

determining the set of mass/charge (m/z) values for all possible contiguous labeled peptide fragments that might result from random cleavages of the peptide backbone;

determining an abundance value from said mass spectrum data for each m/z value in said predetermined set, thereby producing a plurality of abundance values;

calculating a first ranking of the possible sequences for the first label at each residue length by their relative abundances, based on said plurality of abundance values;

calculating a second ranking of the possible sequences for the second label at each residue length by their relative abundances, based on said plurality of abundance values;

calculating a combined ranking for each sequence by linear combination of the first and second rankings at each sequence length; and

calculating a cumulative ranking for the maximum sequence length based on a linear combination of said combined rankings for each residue length of a set of amino acid sequences of the maximum desired sequence length.

82. A method as in claim 81 wherein said mass labels have masses which are different than a mass of each possible amino acid in said set of amino acid sequences.

83. A method as in claim 81 wherein said mass label imparts a unique mass signature to each sequence of said set of amino acid sequences.

84. A method as in claim 81 wherein said labels are different stable isotopes of the same chemical species.

85. A method as in claim 1 wherein said oligomer is a oligosaccharide.

86. A method as in claim 85 wherein said portion of said oligosaccharide is a terminal portion of said oligosaccharide.

87. A method as in claim 86 wherein said terminal portion is a reducing terminus.

88. A method as in claim 86 wherein a label is attached to said portion.
89. A method as in claim 88 wherein said label is covalently bonded to said portion prior to generating said mass spectrum data and wherein said mass spectrum data is transformed from an output of a detector plate.
90. A method as in claim 85 wherein said oligosaccharide is fragmented to generate fragments, comprising said portion, which are then accelerated toward a detector plate to generate said mass spectrum data.
91. A method as in claim 1 wherein said oligomer is a nucleic acid.
92. A method as in claim 91 wherein said portion of said nucleic is a terminal portion of said nucleic acid.
93. A method as in claim 92 wherein said terminal portion is a 3' terminus.
94. A method as in claim 92 wherein a label is attached to said portion.
95. A method as in claim 94 wherein said label is covalently bonded to said portion prior to generating said mass spectrum data and wherein said mass spectrum data is transformed from an output of a detector plate.
96. A method as in claim 91 wherein said nucleic acid is fragmented to generate fragments, comprising said portion, which are then accelerated toward a detector plate to generate said mass spectrum data.

97. A method for processing mass spectrum data to extract specific labeled ions of interest, said method comprising:

determining a substantially periodic block of noise in a mass spectrum data generated from accelerating unlabeled ions to a detector;

filtering said substantially periodic block of noise from said mass spectrum data.

98. A method as in claim 97 wherein the label incorporates one or more elements with an atomic number between 17 and 77, excluding S and P.

99. A method for processing mass spectrum data to extract the relative abundances of two or more differentially labeled ions from different samples, said method comprising:

determining a substantially periodic block of signal in a mass spectrum data generated from accelerating ions labeled with one label mass to a detector;

filtering said substantially periodic block of signal from said mass spectrum data to generate a label 1 filtered mass spectrum;

determining a substantially periodic block of signal in a mass spectrum data generated from accelerating ions labeled with a second label mass to a detector;

filtering said substantially periodic block of signal from said mass spectrum data to generate a label 2 filtered mass spectrum;

100. A method as in claim 99 wherein label 1 and label 2 contain different numbers of mass defect elements with an atomic number between 17 and 77, excluding S and P.

101. A method as in claim 100 wherein the mass defect elements of the label have an atomic number between 35 and 63.

102. A method as in claim 98 wherein the mass defect elements of the label have an atomic number between 35 and 63.

103. A method as in claim 99 wherein the label 1 and label 2 mass spectrum data are compared by ratioing relative abundances of the labeled parent ions, such that the relative amounts of the parent ions can be determined between the two samples.

104. A method as in claim 1 wherein said oligomer is a nucleic acid.

105. A method as in claim 1 wherein said oligomer is an oligosaccharide.

106. A method as in claim 1 wherein said oligomer comprises at least one of a protein, a nucleic acid and a oligosaccharide.

107. A method as in claim 1 wherein said oligomer is labeled prior to being fragmented.

108. A method as in claim 1 where said oligomer is fragmented and the resulting fragments are labeled.

109. A method as in claim 104 wherein said portion is a terminal portion of said nucleic acid.

110. A method as in claim 109 wherein said terminal portion is one of a 3' terminus.

111. A method as in claim 104 wherein said label is covalently bonded to a primer sequence of a nucleic acid prior to the fragments being generated by Sanger, polymerase chain reaction, or Maxam-Gilbert methods and the generation of mass spectrum data.

112. A method as in claim 105 wherein said label is covalently bonded to a reducing terminus of an oligosaccharide prior to enzymatic fragmentation of the oligosaccharide and the generation of mass spectrum data.

113. A method as in claim 6 wherein said protein is fragment by collision-induced-dissociation, either in source or in a collision cell, to generate fragments, comprising said portion, which are then accelerated toward a detector plate to generate said mass spectrum data.

114. A method as in claim 6 wherein said protein is fragmented by partial exoproteolytic digestion prior to generating the mass spectrum data.